IME 775 — Lecture 3

Vectors, Matrices, and Tensors in Machine Learning

---

1. Why Linear Algebra for ML?

At its core, machine learning is about **number crunching**. We need to organize numbers into meaningful structures.

Key Insight: Every input and output in ML can be represented as a vector, matrix, or tensor.

Example:

- An image is a *tensor* of pixel values (height × width × color channels).
- A word embedding is a *vector* of real numbers.
- A dataset is a *matrix* where rows are examples and columns are features.

---

2. Vectors: The Building Blocks

Definition: A vector $\mathbf{x} \in \mathbb{R}^n$ is an ordered sequence of n numbers:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Example (Object Recognition): Output probabilities for {dog, human, cat}:

$$\mathbf{y} = \begin{bmatrix} P(\text{dog}) \\ P(\text{human}) \\ P(\text{cat}) \end{bmatrix}$$

Workout: Write a feature vector for a house with features: [sqft, bedrooms, bathrooms, age]

---

3. Geometric View of Vectors

Key Idea: A vector $\mathbf{x} \in \mathbb{R}^n$ represents a **point** in n-dimensional space.

ML Implication: A model is a geometric transformation mapping input points to output points.

Workout: Sketch vectors $\mathbf{a} = [2, 3]^T$ and $\mathbf{b} = [-1, 2]^T$ in 2D space:

4. Matrices: Data and Transformations

Definition: A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ has m rows and n columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Two Roles (with Examples):

1. **Data Storage:** Each row = one training sample
   Example:

$$\mathbf{A}_{\text{data}} = \begin{bmatrix} 1500 & 3 & 2 & 10 \\ 1800 & 4 & 3 & 5 \\ 1200 & 2 & 1 & 30 \end{bmatrix}$$

   (Each row: [sqft, bedrooms, bathrooms, age] for a house)

2. **Linear Transform:** Maps $\mathbb{R}^n \to \mathbb{R}^m$
Example:

$\mathbf{A}_{\text{transform}}\mathbf{x} = \begin{bmatrix} 2 & 0 \\ \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ \\ x_1 + 3x_2 \end{bmatrix}$ *(Maps a 2D input $\mathbf{x}$ to a 2D output by stretching and mixing the coordinates)*

---

5. Images as Matrices

A grayscale image is a matrix where each entry is pixel brightness (0-255).

$$\mathbf{I} = \begin{bmatrix} 0 & 128 & 255 \\ 64 & 192 & 128 \\ 255 & 64 & 0 \end{bmatrix}$$

- $I_{ij} = 0$: black
- $I_{ij} = 255$: white

Workout: What does $\mathbf{I}^T$ (transpose) do to the image geometrically?

---

6. Tensors: Multidimensional Arrays

Definition: Tensors generalize matrices to arbitrary dimensions.

Object	Dimensions	Example
Scalar	0D	Loss value
Vector	1D	Feature vector
Matrix	2D	Grayscale image
3D Tensor	3D	RGB image
4D Tensor	4D	Batch of images

Example: Batch of 64 RGB images, 224×224:

$$\mathcal{T} \in \mathbb{R}^{64\times3\times224\times224}$$

PyTorch convention: `[batch, channels, height, width]`

---

7. The Dot Product

Definition: For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Workout: Compute $\mathbf{u} \cdot \mathbf{v}$ for $\mathbf{u} = [1, 2, 3]^T$ and $\mathbf{v} = [4, -1, 2]^T$:

---

8. Geometric Meaning of Dot Product

Theorem:

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Proof (Law of Cosines):
Consider the triangle formed by the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$. The side lengths are:

- $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ with included angle θ
- $\|\mathbf{u} - \mathbf{v}\|$ opposite the angle θ

By the Law of Cosines:

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$$

Now expand $\|\mathbf{u} - \mathbf{v}\|^2$ using the dot product:

$$\|\mathbf{u} - \mathbf{v}\|^2 = (\mathbf{u} - \mathbf{v})^T(\mathbf{u} - \mathbf{v}) = \mathbf{u}^T\mathbf{u} - 2\mathbf{u}^T\mathbf{v} + \mathbf{v}^T\mathbf{v} = \|\mathbf{u}\|^2 - 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2$$

Set the two expressions for $\|\mathbf{u} - \mathbf{v}\|^2$ equal and cancel $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ on both sides:

$$-2(\mathbf{u} \cdot \mathbf{v}) = -2\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$$

Divide by -2:

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

> *We form a triangle using $\mathbf{u}, \mathbf{v}$, and $\mathbf{u} - \mathbf{v}$, then apply the Law of Cosines to relate vector lengths to the angle between them.*

<u>Visualisation</u>

Interpretations:

- $\mathbf{u} \cdot \mathbf{v} > 0$: angle $< 90^\circ$ (similar direction)
- $\mathbf{u} \cdot \mathbf{v} = 0$: angle $= 90^\circ$ (orthogonal)
- $\mathbf{u} \cdot \mathbf{v} < 0$: angle $> 90^\circ$ (opposite direction)

Workout: Sketch two vectors with positive, zero, and negative dot products:

---

9. L2 Norm (Euclidean Length)

Definition:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

Properties (with examples):

1. $\|\mathbf{x}\|_2 \geq 0$ (equality iff $\mathbf{x} = \mathbf{0}$)
   Example: For $\mathbf{x} = [3, 4]^T$*,* $\|\mathbf{x}\|_2 = 5 \geq 0$*; for* $\mathbf{x} = [0, 0]^T$*,* $\|\mathbf{x}\|_2 = 0$
2. $\|c\mathbf{x}\|_2 = |c|\|\mathbf{x}\|_2$
   Example: If $c = -2$ *and* $\mathbf{x} = [3, 4]^T$*, then* $\|c\mathbf{x}\|_2 = \| - 2 \cdot [3, 4]^T\|_2 = \|[-6, -8]^T\|_2 = 10 = 2 \times 5 = |c|\|\mathbf{x}\|_2$
3. $\|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$ (Triangle Inequality)
   Example: For $\mathbf{x} = [3, 0]^T, \mathbf{y} = [0, 4]^T$*,* $\|\mathbf{x} + \mathbf{y}\|_2 = \|[3, 4]^T\|_2 = 5 \leq 3 + 4 = 7$

Workout: Compute $\|\mathbf{x}\|_2$ for $\mathbf{x} = [3, 4]^T$:

ML Application: Mean Squared Error = $\frac{1}{n}\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$

---

10. Cosine Similarity

Definition: For non-zero vectors:

$$\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}$$

Example:

$$\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2} = \frac{1 \cdot 1 + 0 \cdot 1}{\sqrt{1^2 + 0^2}\sqrt{1^2 + 1^2}} = \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}$$

where $\mathbf{u} = [1, 0]^T$ and $\mathbf{v} = [1, 1]^T$.

Range: $[-1, 1]$

- $+1$: identical direction
- 0: orthogonal
- -1: opposite direction

Why use cosine similarity? Ignores magnitude — two documents about "ML" are similar even if one is longer.

Workout: Compute cosine similarity for $\mathbf{u} = [1, 0]^T$ and $\mathbf{v} = [1, 1]^T$:

---

11. Orthogonality

Definition: Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** iff:

$$\mathbf{u} \cdot \mathbf{v} = 0 \quad (\text{written } \mathbf{u} \perp \mathbf{v})$$

Why it matters in ML:

- Orthogonal features carry independent information
- PCA produces orthogonal principal components
- Orthogonal weight initialization improves training

Workout: Find a vector orthogonal to $\mathbf{u} = [3, 4]^T$:

Example:

$$\mathbf{v} = [4, -3]^T$$

because:

$$\mathbf{u} \cdot \mathbf{v} = 3 \cdot 4 + 4 \cdot (-3) = 12 - 12 = 0$$

General solution:

$$\mathbf{v} = [-4k, 3k]^T$$

for any scalar k.

12. Matrix Transpose

Definition: For $\mathbf{A} \in \mathbb{R}^{m \times n}$:

$$(\mathbf{A}^T)_{ij} = A_{ji}$$

Properties:

1. $(\mathbf{A}^T)^T = \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
3. $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T$ ← **Reversal!**

Workout: Compute $\mathbf{A}^T$ for $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$:

13. Matrix-Vector Multiplication

Definition: For $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m, \quad y_i = \sum_{j=1}^{n} A_{ij} x_j$$

Interpretation: y_i is the dot product of row i of $\mathbf{A}$ with $\mathbf{x}$.

Workout: Compute $\mathbf{A}\mathbf{x}$ for:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

---

14. Matrix-Matrix Multiplication

Definition: For $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times p}$:

$$\mathbf{C} = \mathbf{A}\mathbf{B} \in \mathbb{R}^{m\times p}, \quad C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Compatibility: $\text{cols}(\mathbf{A}) = \text{rows}(\mathbf{B})$

Critical: $\mathbf{AB} \neq \mathbf{BA}$ in general!

Workout: Verify dimensions: If $\mathbf{A}$ is 3×4 and $\mathbf{B}$ is 4×2, what is the shape of $\mathbf{AB}$?

---

15. Linear Transforms

Definition: $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear iff:

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. $T(c\mathbf{u}) = cT(\mathbf{u})$

Key Theorem: Every linear transform can be written as matrix multiplication:

$$T(\mathbf{x}) = \mathbf{Ax}$$

Geometric Property: Linear transforms preserve collinearity — points on a line map to points on a line.

---

16. Common 2D Transformations

Transform	Matrix	Effect
Identity	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	No change
Scale	$\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$	Stretch/compress
Rotation	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$	Rotate by θ
Shear	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	Slant
Reflection	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	Mirror

Workout: What does $\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$ do to the unit square?

Example:

$$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$

So the matrix stretches the unit square by a factor of 2 in the x-direction and compresses it by a factor of 0.5 in the y-direction.

17. Hyperplanes and Classification

Definition: A hyperplane in $\mathbb{R}^n$ is defined by:

$$\mathbf{w}^T\mathbf{x} + b = 0$$

- $\mathbf{w}$: normal vector (perpendicular to hyperplane)
- b: bias/offset

Dimension intuition:

- In $\mathbb{R}^2$: hyperplane is a line
- In $\mathbb{R}^3$: hyperplane is a plane
- In $\mathbb{R}^n$: it's an (n-1)-dimensional "flat" surface

Why is $\mathbf{w}$ called the "normal" vector? Because it is perpendicular to the hyperplane. Proof: Take any two points $\mathbf{x}_1, \mathbf{x}_2$ that lie on the hyperplane:

$$\mathbf{w}^T\mathbf{x}_1 + b = 0, \quad \mathbf{w}^T\mathbf{x}_2 + b = 0$$

Subtract:

$$\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) = 0$$

This means $\mathbf{w}$ is orthogonal to any direction $(\mathbf{x}_1 - \mathbf{x}_2)$ that lies within the hyperplane. So $\mathbf{w}$ points straight out of the hyperplane.

The hyperplane splits space into two halves: Look at the sign of:

$$s(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

- If $s(\mathbf{x}) > 0$: $\mathbf{x}$ is on one side
- If $s(\mathbf{x}) < 0$: $\mathbf{x}$ is on the other side
- If $s(\mathbf{x}) = 0$: $\mathbf{x}$ is on the boundary

That's why it's a decision boundary in classification.

Linear Classifier:

$$\hat{y} = \mathrm{sign}(\mathbf{w}^T\mathbf{x} + b)$$

Workout: For $\mathbf{w} = [1, 2]^T$ and $b = -3$, sketch the decision boundary:

solution:

$$\mathbf{w}^T\mathbf{x} + b = 0$$

$$1x + 2y + (-3) = 0$$

$$x + 2y - 3 = 0$$

$$y = \frac{3 - x}{2}$$

18. Parametric Line Equation

Theorem: Any point on the line through $\mathbf{p}$ and $\mathbf{q}$:

$$\mathbf{r}(\alpha) = \alpha\mathbf{p} + (1 - \alpha)\mathbf{q}$$

Interpretation:

- $\alpha \in [0, 1]$: between $\mathbf{p}$ and $\mathbf{q}$
- $\alpha = 0.5$: midpoint
- $\alpha < 0$ or $\alpha > 1$: outside segment

Workout: Find the midpoint of $\mathbf{p} = [1, 2]^T$ and $\mathbf{q} = [5, 6]^T$:

solution:

$$\mathbf{r}(\alpha) = \alpha\mathbf{p} + (1 - \alpha)\mathbf{q}$$

$$\mathbf{r}(0.5) = 0.5\mathbf{p} + (1 - 0.5)\mathbf{q}$$

$$\mathbf{r}(0.5) = 0.5[1, 2]^T + 0.5[5, 6]^T$$

$$\mathbf{r}(0.5) = [0.5, 1] + [2.5, 3]$$

$$\mathbf{r}(0.5) = [3, 4]$$

19. Linear Independence

Definition: Vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are **linearly independent** iff:

$$\sum_{i=1}^{k} \alpha_i \mathbf{v}_i = \mathbf{0} \implies \alpha_i = 0 \text{ for all } i$$

Intuition: No vector can be written as a combination of the others.

Workout: Are $\mathbf{v}_1 = [1, 0]^T$, $\mathbf{v}_2 = [0, 1]^T$, $\mathbf{v}_3 = [1, 1]^T$ linearly independent?

solution:

$$\sum_{i=1}^{3} \alpha_i \mathbf{v}_i = \mathbf{0}$$

$$\alpha_1[1, 0]^T + \alpha_2[0, 1]^T + \alpha_3[1, 1]^T = [0, 0]^T$$

$$\alpha_1 + \alpha_3 = 0$$

$$\alpha_2 + \alpha_3 = 0$$

$$\alpha_1 = -\alpha_3$$

$$\alpha_2 = -\alpha_3$$

$$\alpha_3 = 0$$

---

20. Span and Basis

Definition (Span): All possible linear combinations:

$$\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \left\{ \sum_i \alpha_i \mathbf{v}_i : \alpha_i \in \mathbb{R} \right\}$$

Definition (Basis): A set of vectors is a **basis** for $\mathbb{R}^n$ if:

1. Linearly independent
2. Spans $\mathbb{R}^n$

Standard Basis for $\mathbb{R}^3$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Summary: Key Formulas

Concept	Formula
Dot Product	$\mathbf{u} \cdot \mathbf{v} = \sum_i u_i v_i$
L2 Norm	$\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$
Cosine Similarity	$\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$
Orthogonality	$\mathbf{u} \cdot \mathbf{v} = 0$
Matrix-Vector	$y_i = \sum_j A_{ij} x_j$
Hyperplane	$\mathbf{w}^T \mathbf{x} + b = 0$

ML is geometry in high dimensions: data live in spans, models choose bases, and classifiers cut space with hyperplanes.